

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 9, 2002, 17:57:04 : Search time 1595 Seconds
(without alignments)
6811.943 Million cell updates/sec

Title: US-09-635-501-2
Perfect score: 4291
Sequence: 1 M5SSSWLLSLVTAQAQST.....ISKGENNPGQNTDDVQTSF 805

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US09635501/tunat_09102002_094529_18537/app_query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcg -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09635501@cgn.1.1106 -runat_09102002_094529_18537 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAB -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1452	33.8	987	9	AL551235
2	1071	25.0	1603	11	AK008530 Mus muscu

3	1054.5	24.6	763	10	BI913504	BI913504	603179133
4	1048.5	24.4	879	10	EG401683	EG401683	602466350
5	1031	24.0	766	10	BG962298	BG962298	602827090
6	1022	23.8	676	10	BI561069	BI561069	60254275
7	979	22.8	635	10	BG722079	BG722079	602698682
8	901	21.0	609	10	BI561359	BI561359	60255121
9	864	20.1	965	10	BF789159	BF789159	602105037
10	857	20.0	555	9	AA162058	AA162058	ms31h11.r
11	819	19.1	686	9	BG652968	BG652968	BB652968
12	781	18.2	452	10	EG48060	EG48060	602501471
13	749.5	17.5	556	10	BM030353	BM030353	488973 MA
14	730.5	17.0	741	10	BG772384	BG772384	602722306
15	687.5	16.0	800	10	BI26471	BI26471	603076177
16	684	15.9	1154	10	BM460886	BM460886	AGENCOCURT
17	680	15.8	900	10	BI415293	BI415293	602987289
18	672.5	15.7	778	10	BI831019	BI831019	603080858
19	661.5	15.4	848	10	BI904144	BI904144	603166751
20	659.5	15.4	842	10	BF784147	BF784147	602108008
21	646	15.1	848	10	BI766340	BI766340	603052444
22	643	15.0	492	9	AW260204	AW260204	um74 f04.x
c	625	14.6	348	9	AL039738	AL039738	DKEZP434M
23	622	14.5	773	10	BI830936	BI830936	603081067
24	622	14.5	773	10	BI830936	BI830936	603081067
25	615	14.3	887	10	BI196733	BI196733	602755360
c	593	13.8	347	9	AW848579	AW848579	IL3-C7021
27	593	13.8	666	10	BI686319	BI686319	603333252
28	588.5	13.7	725	10	BI821905	BI821905	603039709
29	578	13.5	751	10	BI828973	BI828973	603074849
30	575.5	13.4	471	9	AW358360	AW358360	42453 MAR
31	564.5	13.2	823	10	BI914635	BI914635	603179402
32	564	13.1	606	9	AW322773	AW322773	u052b11.y
c	553	12.9	728	9	AW026259	AW026259	wv10g09.y
34	549	12.8	459	9	BE138336	BE138336	ug51d09.y
c	549	12.8	574	9	AI116021	AI116021	ue25h10.x
36	548	12.8	432	9	AW258861	AW258861	um74 f04.y
37	547	12.7	865	10	BI825221	BI825221	603071908
38	544	12.7	965	10	BI453005	BI453005	603170164
39	543.5	12.7	775	10	BI103238	BI103238	602899478
40	542	12.6	628	10	BM487251	BM487251	p6m2n.pk0
41	539	12.6	418	9	BB687602	BB687602	BB687602
c	538	12.5	749	9	AW026258	AW026258	wv10g07.x
43	535	12.5	896	10	BF783379	BF783379	602110217
44	534.5	12.5	969	10	BF780858	BF780858	602105748
c	533.5	12.4	755	9	AI971658	AI971658	wr06g04.x

ALIGNMENTS

RESULT 1	AL551235	987 bp	mRNA	linear	EST 16-FEB-2001
LOCUS	AL551235	LTI_NFL006_PL2	Homo sapiens	CDNA clone	CS0D1041YM14 5
DEFINITION	AL551235	prime, mRNA sequence.			
ACCESSION	AL551235				
VERSION	AL551235.1	GI:12888986			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 987)				
JOURNAL	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.				
COMMENT	Full-length cDNA libraries and normalization				
Unpublished (2001)	Contact: Genoscope				
Genoscope - Centre National de Sequencage	BP 191 91006 EVRY cedex - France				
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.	Location/Qualifiers				
FEATURES	1..987				
Source	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0D1041YM14"				
	/clone_lib="LTI_NFL006_PL2"				

```
/tissue_type="placenta"
/notes=vector: pcwvSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pcwvSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com
BASE COUNT 322 a 182 c 211 g 270 t 2 others
ORIGIN
```

```
Alignment Scores:
Pred. No.: 3.74e-156 Length: 987
Score: 1452.00 Matches: 278
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.84% Indels: 0
DB: 9 Gaps: 0
```

US-09-635-501-2 (1-805) x AL551235 (1-987)

```
QY 528 AlaLeuGluGlnAlaAlaLysHisGluGlyProLeuHisLysCysAspIleSerAnSer 547
DB 3 GCACCTTTCAGCAGCAGTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAACTCT 62
QY 548 ThrGluAlaGlnLysLeuPheAsnMetLeuArgLeuGlyLysSerGluProTTPThr 567
DB 63 ACAGAGCTGGACAGAACTGTTCAATATGCTGAGGCTTGAATAATCAGAACTCCCTGGACC 122
QY 568 LeuAlaLeuGluGlnValGlyAlaLysAsnMetAsnValArgProLeuLeuAsnTyr 587
DB 123 CTAGCATTTGGAAATCTTGTAGGAGCAAGAACATGAATGTAAGGCCACCTGCCTCAACTAC 182
QY 588 PheGluProLeuPheThrTrpLeuLysAspGlnAsnLysAsnSerPheValGlyTrpSer 607
DB 183 TTTGAGCCCTTATTACCTGGCTGAAAGACCAAGCAAGAAATCTTTTGTGGATGGAGT 242
QY 608 ThrAspTrpSerProTyrAlaAspGlnSerIleLysValArgIleSerLeuLysSerAla 627
DB 243 ACCGACTGGAGTCCATATGCAGACCAAGCAAGCATCAAGTGAAGTAAAGCTTAAATCAGCT 302
QY 628 LeuGlyAspLysAlaTyrGluTrpAsnAspAsnGluMetTyrLeuPheArgSerVal 647
DB 303 CTTGGAGATAAGCATATGAATGAACGACAAATGAATGTAATGTAATGTAATGTAATGTAAT 362
QY 648 AlaTyrAlaMetArgGlnTyrPheLeuLysValLysAsnGlnMetIleLeuPheGlyGlu 667
DB 363 GCATATGCTATGAGCAGTACTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 422
QY 668 GluAspValArgValAlaAsnLeuLysProArgIleSerPheAsnPheValThrAla 687
DB 423 GAGGATGTGGAGTGGCTTAATTTGAAACCAAGATCTCTTAAATTTCTTTGTGCACTGCA 482
QY 688 ProLysAsnValSerAspIleProArgThrGluValGluLysAlaIleArgMetSer 707
DB 483 CCTAAATATGTCTGATATCATCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 542
QY 708 ArgSerArgIleAsnAspAlaPheArgLeuAsnAspAsnSerLeuGluPheLeuGlyIle 727
DB 543 CGGAGCCGTATCAATGATGCTTTCCGCTGAATGACACACGCTAGAGTTCTTGGGGATA 602
QY 728 GlnProThrLeuGlyProProAsnGlnProValSerIleTrpLeuIleValPheGly 747
DB 603 CAGCCACACCTTGGACCTCTCAACACCCCTGTTTCCATATGCGTGATGTTTGGG 662
QY 748 ValValMetGlyValIleValValGlyIleValIleLeuIlePheThrGlyIleArgAsp 767
DB 663 GTTGTGATGGAGTATAGTGGTGGCATTTGTCATCTGATCTCTCACTGGATCAGAGAT 722
QY 768 ArgLysLysLysAsnLysAlaArgSerGlyGluAsnProTyrAlaSerIleAspIleSer 787
```

```
DB 723 CGGAGAGAGAAAATAAAGCAAGAGTGGAGAAATCCTTATGCTCCATCGATATAGC 782
QY 788 LysGlyGluAsnAsnProGlyPheGlnAsnThrAspValGlnThrSerPhe 805
DB 793 AAAGAGAGAAATAATCCAGGATTCACAAACACTGATGATGTCAGACCTCTTT 836
RESULT 2
AK008530 1603 bp mRNA linear HTC 19-JAN-2002
LOCUS Mus musculus adult male small intestine cDNA, RIKEN full-length
DEFINITION enriched library, clone:2010305L05:homolog to ANGIOTENSIN
CONVERTING ENZYME-LIKE PROTEIN (ACE-RELATED CARBOXYPEPTIDACE2),
full insert sequence.
ACCESSION AK008530
VERSION AK008530.1 GI:12842766
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2010305L05.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Alzawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1603)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Balgarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Schroth,L., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Seitama,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamanaka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
```

Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

FEATURES

Location/Qualifiers

1..1603
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGI:1902242"
/db_xref="taxon:10090"
/clone="2010305L05"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
49..798
/note="data source: SPTR, source key: Q9NRA7, evidence: ISS
homolog to ANGIOTENSIN CONVERTING ENZYME-LIKE PROTEIN
(ACE-RELATED CARBOXYPEPTIDASE ACE2)
putative"

CDS

49..798

/note="data source: SPTR, source key: Q9NRA7, evidence: ISS
homolog to ANGIOTENSIN CONVERTING ENZYME-LIKE PROTEIN
(ACE-RELATED CARBOXYPEPTIDASE ACE2)
putative"
/codon_start=1
/protein_id="BAB25723.1"
/db_xref="GI:12842767"
/translation="MLSLGNSEPTKALENVGARNMDVKPLNLTQFLEWLEQNR
NSFQWNTENSPYADQSIKVRISKALGANAYENNMFLRSVAYAMKRFYSII
KNQTFPEEDVRVSDLRKPRVSFFVTSPQNSVDIPRSEVEDAIRMGRGRINDVFG
LNDNSLEFLGHPTLEPPYPPPTIWIIFGVYMALVWVGIIILIVTGIGRKKNET
KRENSYDSMDIGKSGESNAGSQNSDDAQTSF"
1584..1589
polyA_signal
/note="putative"
1603
polyA_site
/note="putative"

BASE COUNT 465 a 321 c 366 g 450 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 5.69e-112 Length: 1603
Score: 1071.00 Matches: 200
Percent Similarity: 87.50% Conservative: 31
Best Local Similarity: 75.76% Mismatches: 33
Query Match: 24.96% Indels: 0
DB: 11 Gaps: 0

us-09-635-501-2 (1-805) x AK008530 (1-1603)

QY 542 CysAspIleSerAnsSerThrGluAlaGlyGlnLysLeuPheAsnMetLeuArgLeuGly 561
|||||
Db 4 TGTGACATCTCAATTCCTGAACTGGGCGAGAGTGTCTCAAGATGCTGAGTCTTGGG 63
QY 562 LysSerGluProTrpThrLeuAlaLeuGluAsnValValGlyAlaLysAsnMetAsnVal 581
|||||
Db 64 AATTACAGACCTGGACCAAGCCTTGGAAATGTGGTAGGAGCAAGAAATATGATGTA 123
QY 582 ArgProLeuLeuAsnThrPheGluProLeuPheThrTrpLeuLysAspGlnAsnLysAsn 601
|||||
Db 124 AAACCATGCTCAATTACTTCCAAACCGTGTGTGACTGGCTGAAAGAGACAGAAAT 183
QY 602 SerPheValGlyTrpSerThrAspTrpSerProTyrAlaAspGlnSerIleLysValArg 621
|||||
Db 184 TCCTTTGGGGTGAACACTGAATGGAGCCCATATGCCAGCAACCAATTAAGTCAGG 243
QY 622 IleSerLeuLysSerAlaLeuGlyAspLysAlaTyrGluTrpAsnAspAsnGluMetTyr 641
|||||
Db 244 ATAAGCCTAAATACAGCTCTTGGAGCTAATGATATGATGAGCAACAGAAATGTC 303
QY 642 LeuPheArgSerValAlaTyrAlaMetArgGlnTyrPheLeuLysValLysAsnGln 661
|||||
Db 304 CTGTTCCGATCATCTGTTCATATGCCATGAGAGAAATATTTTCAATTAATCAAAACAG 363
QY 662 MetIleLeuPheGlyGluGluAspValArgValAlaAsnLeuLysProArgIleSerPhe 681

Db 364 ACAGTTCCTTTCTAGAGGAGAGTACGAGTGACGTGATTTGAACCAAGAGTCTCCTTC 423
QY 682 AsnPhePheValThrAlaProLysAsnValSerAspIleIleProArgThrGluValGlu 701
Db 424 TACTTCTTTGTACCTCACCCCAAAATGTCTGTATGTCTATCTCTAGAGTGAAGTTGAA 483
QY 702 LysAlaIleArgMetSerArgSerArgIleAsnAspAlaPheArgLeuAsnAspAsnSer 721
Db 484 GATCCATCAGGATGTCGGGGCCGCATCAATGATGTCTTTGGCCTCAATGATAACAGC 543
QY 722 LeuGluPheLeuGlyIleGlnProThrLeuGlyProProAsnGlnProProValSerIle 741
Db 544 CTGGAGTTCTGGGATTCACCAACACTTGAGCCACCTTACCCAGCCTCTCTGTCCACATA 603
QY 742 TrpLeuIleValPheGlyValValMetGlyValIleValValGlyIleValIleLeuIle 761
Db 604 TGGCTGATTATTTTGTGTGTGTGATGGCACTGGTAGTGGTGGCATCATCTCGATT 663
QY 762 PheThrGlyIleArgAspArgLysLysLysAsnLysAlaArgSerGlyGluAsnProTyr 781
Db 664 GTCACTGGGATCAAGGTCGAAAGAGAAATAAGAAAGAGAGAGAGAGAACTCTTAT 723
QY 782 AlaSerIleAspIleSerLysGlyGluAsnAsnProGlyPheGlnAsnThrAspAspVal 801
Db 724 GACTCGATGACATTCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 783
QY 802 GlnThrSerPhe 805
Db 784 CAGACTTCCTTT 795

RESULT 3

BI913504

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI913504 763 bp mRNA linear EST 16-OCT-2001
603179133F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5243048 5',
mRNA sequence.

EST.
BI913504.1 GI:16177911
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 763)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.
Email: rstra@nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLML1613 row: b column: 09
High quality sequence stop: 675.
Location/Qualifiers

1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5243048"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library.*

```
BASE COUNT      237 a  167 c  170 g  189 t
ORIGIN
Alignment Scores:
Pred. No.:      1,54e-110      Length:      763
Score:          1054.50      Matches:      225
Percent Similarity: 91.30%      Conservative: 6
Best Local Similarity: 88.93%      Mismatches: 12
Query Match:      24.57%      Indels:      10
DB:              10      Gaps:      3

US-09-635-501-2 (1-805) x BI913504 (1-763)
Qy 1 MetSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
Db 19 ATGTCAAGCTCTTCCTGGCTCTCTCAGCCTTGTGTGAATGCTGCTCAGTCCACC 78
Qy 21 Ile-GluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPh 40
Db 79 ATGTGAGACACGCCCAAGACATTTTGGACAAGTTTAAACCACGAGCCGAGACCTGT 138
Qy 40 eTyGlnSerSerLeuAlaSerTrpAsnTrpAsnThrAsnThrGluGluAsnValG1 60
Db 139 CTATCAAGTTCATCTGCTTCTTGGAAATTAACACCAATATTACTGAAGAAGATGTCCA 198
Qy 60 nAsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAl 80
Db 199 AAACATGAATATGCTGGGACAAATGCTGCTCCCTTTTAAAGGACAGTCCACACTGC 258
Qy 80 aGlnMetTrpProLeuGlnGlnLeuGlnAsnLeuThrValLysLeuGlnAlaLe 100
Db 259 CCAAAATGTATCCACTACAAGAAATTCAGAAATTCACAGTCAAGCTTCAGCTGCAGGCTCT 318
Qy 100 uGlnGlnAsnGlySerSerValLeuSerGluAspLysSerLysArgLeuAsnThrIleLe 120
Db 319 TCACCAAAATGGGTCTCAGTGTCTCAGAACAGACGAAACGGTTGAACAAATCT 378
Qy 120 uAsnThrMetSerThrIleTySerThrGlyLysValCysAsnProAsnProGlnG1 140
Db 379 AAATACAATGAGCACCATCTACAGTACTGGAAGTTTGTAAACCAGATAATCCACAAGA 438
Qy 140 uCysLeuLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyAsnG1 160
Db 439 ATGCTTATTTACTTGAACCAAGTTTGACTGACATATGGCAACAGCTTTAGACTACAATGA 498
Qy 160 uArgLeuTrpAlaTrp---GluSerTrpArgSerGluValGlyLysGlnLeuArgProLe 179
Db 499 GAGGCTCTGGGCTTGGCAAAACGCTGGCAGATCTGAGGTGCGCAAGCAGCTGAGGCCATT 558
Qy 179 uTyTrGluGluTyValValLeu-LysAsnGluMetAlaArgAlaAsnHisTyTrGluAspT 199
Db 559 ATATGAAGATGTGTGCTTGTGAACATGAGATGGCAAGCAATCATTTATGAGGACT 618
Qy 199 yRGlyAspTyTrpArg-GlyAspTyRGluVal-AsnGlyVal-AspGlyTyTrpAspTyrs 218
Db 619 ATGGGATTTATCGGAGAAGGAGACTATGAAGTACAATGGGTAGCATGGCTATGACTACC 678
Qy 218 eArgGly-GlnLeuIleGlu---AspValGluHisThrPhe---GluGluIleIlePro 235
Db 679 GCCCGGACGAGTTGATCGAAGCTGTGGAAACATACCCTTTGACGAGACTTTAAACCCA 738
Qy 236 LeuTyTr-GluHisLeuHisAlaTyTr 243
Db 739 TTATACTGAACCTCTTCATGCTAT 763

RESULT 4
BG401683
LOCUS      BG401683      879 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION 602466350F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4594140 5',
            mRNA sequence.
ACCESSION  BG401683
```

```
VERSION      BG401683.1      GI:13295131
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 879)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cga@bbs-r@mail.nih.gov
  Tissue Procurement: CLONTECH Laboratories, Inc.
  cDNA Library Preparation: CLONTECH Laboratories, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LHCMI335 row: d column: 13
  High quality sequence stop: 619.
  Location/Qualifiers
    1..879
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4594140"
      /lab_host="NIH_MGC_75"
      /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
      SfiI (ggccgctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
      3' adaptors were used in cloning as follows: 5' adaptor
      sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor sequence:
      5'-ATTCTAGAGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
      C, or G and N = A, C, G, or T). Average insert size 1.65
      kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
      by PCR. This library was enriched for full-length clones
      and was constructed by Clontech Laboratories (Palo Alto,
      CA). Note: this is a NIH_MGC Library."
BASE COUNT  275 a  177 c  210 g  216 t  1 others
ORIGIN
Alignment Scores:
Pred. No.:      9.29e-110      Length:      879
Score:          1048.50      Matches:      218
Percent Similarity: 80.50%      Conservative: 9
Best Local Similarity: 77.30%      Mismatches: 12
Query Match:      24.43%      Indels:      45
DB:              10      Gaps:      4

US-09-635-501-2 (1-805) x BG401683 (1-879)
Qy 388 GlnProPheLeuLeuArgAsnGlyAlaAsnGluGlyPheHisGluAlaValGlyGluIle 407
Db 2 CAACCTTTCTGCTTGAAGAAATGAGAGCTTAATGAAGGATCCATGAAGCTGTGGGGAATC 61
Qy 408 MetSerLeuSerAlaAlaThrProLysHisLeuLysSerIleGlyLeuLeuSerProAsp 427
Db 62 ATGTCACCTTCTGCGAGCCACACCTAAAGCATTTAAATCCATTTGCTCTCTGTCCACCCGAT 121
Qy 428 PheGlnGluAspAsnGluThrGluIleAsnPheLeuLysGlnAlaLeuThrIleVal 447
Db 132 TTTCAAGAAGACAAATGAACAGAAATAAATCTCCTGCTCAAAACAGCCTACGATTGT- 180
Qy 448 GlyThrLeuProPheThrTyTrMetLeuGluLysTrpArgTrpMetValPheLysGlyGlu 467
Db 181 GGGACTCTGCCATTTTACTTACATGTTAGAGAACTGGAGGTGGATGCTCTTTAAAGGGAA 240
Qy 468 IleProLysAspGlnTrpMetLysLysTrpTrpGluMetLysArgGluIleValGlyVal 487
Db 241 ATTCCTCAAGACCAAGTGGATGAAAGAGTGGTGGAGATGAAG----- 282
Qy 488 ValGluProValProHisAspGluThrTyTrCysAspProAlaSerLeuPheHisValSer 507
```

Db 282 ----- 282
 QY 508 AsnAspTyrSerPheIleArgTyrThrArgThrLeuTyrGlnPheGlnPheGlnGlu 527
 Db 283 -----ACCCCTTACCAATTCAGTTTCAGAA 309
 QY 528 AlaLeuGlnAlaAlaLysHisGluGlyProLeuHisLysCysAspIleSerAsnSer 547
 Db 310 GCACCTTTCAGCAGCAGTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAACTCT 369
 QY 548 ThrGluAlaGlyGlnLysLeuPheAsnMetLeuArgLeuGlyLysSerGluProThrPhe 567
 Db 370 ACAGAGCTGTGACAGAACCTTCAATATGCTGAGGCTTGGAAATTCAGAACCTTGGACC 429
 QY 568 LeuAlaLeuGluAsnValValGlyAlaLysAsnMetAsnValArgProLeuLeuAsnTyr 587
 Db 430 CTAGCATTTGGAATTTGTAGGAGCAAGACATGAATGTAAAGCCACTGCTCAACTAC 489
 QY 588 PheGluProLeuPheThrTrpLeuLysAspGluAsnLysAsnSerPheValGlyTrpSer 607
 Db 490 TTTGAGCCCTTATTATCTGCTGGCTGAAGACCAAGAAATTCCTTTTGTGGGATGGAGT 549
 QY 608 ThrAspTrpSerProTyrAlaAspGlnSerIleLysValArgIleSerLeuLysSerAl 627
 Db 550 ACCGACTGAGTCTCATATGAGACCAACATCAAGTGAAGTAAAGCTAAATCAGTC 609
 QY 627 aLeuGlyAspLysAlaTyrGluTrpAsnAspAsnGluMetTyrLeuPheArgSerSerVa 647
 Db 610 T----GGAGATAGCA-TATGACTGGAGC---ACATGAATGTCTCTNCCGATCATCTGT 662
 QY 647 AlaTyr-AlaMetArgGlnTyrPheLeuLysValLysAsnGlnMetIleLeuPheGlyG 667
 Db 663 GGCATATTGCTATTACGGAGTACTTTTAAAGTAAACATCG-----ATGATCTTTGGGG 716
 QY 667 lu 667
 Db 717 AG 718

RESULT 5
 LOCUS BG962298
 DEFINITION 602827090F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4981606 5',
 mRNA sequence.
 ACCESSION BG962298
 VERSION BG962298.1 GI:14349935
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10983 row: d column: 23
 High quality sequence stop: 708.
 Location/Qualifiers
 1. .766
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4981606"
 /clone_lib="NCI_CGAP_Co24"
 /lab_host="DH10B (T1 phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally Primer: Oligo dt.
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 233 a 156 c 179 g 198 t
 ORIGIN

Alignment Scores:

Pred. No.: 7,89e-108 Length: 766
 Score: 1031.00 Matches: 210
 Percent Similarity: 88.85% Conservative: 21
 Best Local Similarity: 80.77% Mismatches: 23
 Query Match: 24.03% Indels: 9
 DB: 10 Gaps: 0

US-09-635-501-2 (1-805) x BG962298 (1-766)

QY 416 LysHisLeuLysSerIleGlyLeuLeuSerProAspPheGlnGluAspAsnGluThrGlu 435
 Db 2 AAGCATCTGAATCCATTTGGTCTTTCGCATCCGATTTTCAAGAAGATAGCAAGACAGAG 61
 QY 436 IleAsnPheLeuLeuLysGlnAlaLeuThrIleValGlyThrLeuProPheThrTrpMet 455
 Db 62 ATAACTTCTCTACTGAAACAGGCATTTGCAATTTGTTGAACACTACCGTTTACTTACATG 121
 QY 456 LeuGluLysTrpArgTrpMetValPheLysGlyGluIleProLysAspGlnTrpMetLys 475
 Db 122 TTAGAGAACTGGAGGTGATGCTCTTTCGGGGTGAATTTCCAAAGAGCAGTGGATGAA 181
 QY 476 LysTrpTrpGluMetLysArgGluIleValGlyValValGluProValProHisAspGlu 495
 Db 182 AAGTGGTGGAGATGAACGGGAGATCGTTGGTGTGTGGAGCCTCTGCGCTCATGATGAA 241
 QY 496 ThrTyrCysAspProAlaSerLysPheHisValSerAsnAspTyrSerPheIleArgTyr 515
 Db 242 ACATACTGTGACCTGCATCTCTCTCCATGTTTCTAATGATTACTCATTCATTCGATAT 301
 QY 516 TyrThrArgThrLeuTyrGlnPheGlnPheGlnAlaLeuCysGlnAlaAlaLysHis 535
 Db 302 TACACAAGAGCATTATACCAATTCAGTTTCCAAAGCTCTTTGTCAAGCAGCTAAGTAT 361
 QY 536 GluGlyProLeuHisLysCysAspIleSerAsnSerThrGluAlaGlyGlnLysLeuPhe 555
 Db 362 AATGGTCTCTGCACAAATGTGACATCTCAATTCACACTGAGCTGGCAGAACTGTCTC 421
 QY 556 AsnMetLeuArgLeuGlyLysSerGluProThrTrpThrLeuAlaLeuGluAsnValGlu 575
 Db 422 AAGATGCTGAGTCTTTGGAAATTCAGAGCCCTGGACCAAGCCTTGGAAACATGTGTAGG 481
 QY 575 ValAlaLysAsnMetAsnValArgProLeuLeuAsnTyrPheGluProLeuPheThrTrpLe 595
 Db 482 ACAAGGAATATGGATGTAAACCCACTGCTGCTCAATTTACTTCCACCGTGTGTGACGTGG 541
 QY 595 uLysAspGlnAsnLysAsnSerPhe-ValGly-TrpSerThrAspTrpSerProTyrAla 614
 Db 542 GAAAGAGCAGACAGAAATTTCTTTCAGTGGCGTGGAGCCTGGAACACTGCAATGGAGCCCATATGCC 601
 QY 615 AspGlnSerIleLysValArgIle-SerLeuLysSerAlaLeuGlyAsp-LysAlaTyrC 634
 Db 602 GACCAAGCATTTACAGTGAAGTAAAGCCTAAGCTAGCTCTTGGAGCTACATGTCATATG 661
 QY 634 LuTrpAsnAsp-AsnGluMetTyrLeuPheArgSerSerValAlaTyrAlaMetArgGln 653
 Db 662 ACTGGACCAACGACGAATTTCTCTGTTCCGATCATCTGT-GCATATGCCATGAGAAC- 719
 QY 654 TyrPheLeuLysValLysAsnGlnMetIleLeuPheGlyGluGluAsp 669
 Db 720 GTCATGCTCAATAATCAAAAACAGACAGCTCTCTTTTCTA-GAGGAGGAT 766

RESULT 6

BI561069

LOCUS BI561069

DEFINITION 603254275F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296531 5',

676 bp mRNA linear EST 05-SEP-2001

mRNA sequence.
 ACCESSION BI561069
 VERSION BI561069.1 GI:15448383
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 676)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM11750 row: f column: 20
 High quality sequence stop: 674.
 Location/Qualifiers
 1..676
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5298531"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 215 a 134 c 161 g 166 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.16e-107 Length: 676
 Score: 1022.00 Matches: 205
 Percent Similarity: 97.62% Conservative: 0
 Best Local Similarity: 97.62% Mismatches: 1
 Query Match: 23.82% Indels: 4
 DB: 10 Gaps: 0

US-09-635-501-2 (1-805) x BI561069 (1-676)

QY 1 MetSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
 Db 55 ATGTCAAGCTCTTCTGCTCTCTCAGCTTGTGTGTAACCTGCTCAGTCCACC 114
 QY 21 ILeGluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPhe 40
 Db 115 ATTGAGGACAGCCCAAGCATTTGGGACAACTTTAACCCAGAACGCAAGACCTGTTC 174
 QY 41 TyrGlnSerSerLeuAlaSerTrpAsnTyrAsnThrAsnIleThrGluGluAsnValGln 60
 Db 175 TATCAAAAGTTCACTTGCTTCTTGAATATTATACCAATATTACTGAAGAAATGTCAA 234
 QY 61 AsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGlu-GlnSerThrLeuAl 80
 Db 235 AACATGATATGCTGGGGCAAAATGGTCTGCTTTTAAAGGAAGACAGTCCACACTGC 294
 QY 80 acInMetTyrProLeuGlnGluIleGlnAsnLeuThrValLysLeuGlnAlaLe 100
 Db 295 CCAATGTATCCACTACAGAAATTCAGAAATCTCAGTCAGTTCAGCTCAGGCTCT 354

QY 100 uClnGlnAsnGlySerSerValLeuSerGluAspLysSerLysArgLeuAsnThrIleLe 120
 Db 355 TCAGCAAAATGGGTCTTCACTGCTCTCAGAAGACAAAGCAAAACGGTTGAACACAATCT 414
 QY 120 uAsnThrMetSerThrIleTyrSerThrGlyLysValCysAsnProAspAspProGlnG 140
 Db 415 AAATACAAATGAGCACCACATACAGTACTGGAAGAAGTTTGTAAACCCAGATAATCCACAAGA 474
 QY 140 uCysLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyrAsnG 160
 Db 475 ATGCTTATTACTTCAACACAGGTTTGAATGAATAATGCGCAACAGTTTAGACTACAATGA 534
 QY 160 uArgLeuTrpAlaTrpGluSerTrpArgSerGluValGlyLysGlnLeuAtgProLeuTy 180
 Db 535 GAGGCTCTGGGCTTGGGAAAGCTGGAGATCTGAGGTGCGGACAGCTGAGGCCATTATA 594
 QY 180 rGluGluTyrValValLeu-LysAsnGluMetAlaArgAlaAsnHistyrGlu-AspTyr 199
 Db 595 TGAAGAGTATGTGCTCTTGAATAAATGAGATGCAAGACAAATCATTTATGAGGACTAT 654
 QY 200 -GlyAspTyrTrpArgGlyAsp 206
 Db 655 GGGGATTTATGGAGAGGAGAC 676

RESULT 7
 BG722079 635 bp mRNA linear EST 08-MAY-2001
 LOCUS 602698682F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4830668 5',
 DEFINITION mRNA sequence.
 ACCESSION BG722079
 VERSION BG722079.1 GI:14001266
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 635)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM10752 row: c column: 21
 High quality sequence stop: 633.
 Location/Qualifiers
 1..635
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4830668"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 202 a 131 c 144 g 158 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5.85e-102 Length: 635

Score: 979.00 Matches: 197
 Percent Similarity: 98.99% Conservativity: 0
 Best Local Similarity: 98.99% Mismatches: 2
 Query Match: 22.82% Indels: 2
 DB: 10 Gaps: 0

US-09-635-501-2 (1-805) x BG722079 (1-635)

QY 1 MetSerSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
 Db 41 ATGTCAAGCTCTCTGCTGCTCTCTCAGCCCTGT-GCTGTAACTGCTCAGTCCACC 99
 QY 21 IleGluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPhe 40
 Db 100 ATTGAGAACAGGCCAAGACATTTTG-GACAAGTTTAAACACGAGCGGAGACCTGTC 158
 QY 41 TyrGlnSerSerLeuAlaSerTrpAsnTrpAsnThrAsnIleThrGluGluAsnValGln 60
 Db 159 TATCAAGTTCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 218
 QY 61 AsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAla 80
 Db 219 AACATGAATTAATGCTGGGACAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 278
 QY 81 GlnMetTrpProLeuGlnGluIleGlnAsnLeuThrValLysLeuGlnAlaLeu 100
 Db 279 CAANTGTATCCACTACAGAATTCAGAAATCTCAGATCAAGTTCAGTCCAGCTCTT 338
 QY 101 GlnAsnGlnSerSerValLeuSerGluAspLysSerLysArgLeuAsnThrIleLeu 120
 Db 339 CAGCAAAATGGGTCTCTGCTGCTCAGAACAGCAAGCAACGGTTGACCAATCTA 398
 QY 121 AsnThrMetSerThrIleTrpSerThrGlyLysValCysAsnProAspAsnProGlnGlu 140
 Db 399 AATCAATGAGCACCACATCTACAGTACTGAAAGTTGTAAACAGATTAATCCAGAA 458
 QY 141 CysLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyrAsnGlu 160
 Db 459 TGTATTATTGTAACCAAGTTTGAATGAATGAATGAATGAATGAATGAATGAATGAAT 518
 QY 161 ArgLeuTrpAlaTrpGluSerTrpArgSerGluValGlyLysGlnLeuArgProLeuTyr 180
 Db 519 AGCTCTGGCTGGGAAAGCTGGAGATCTGAGTGGCAAGCAGCTGAGGCGCATATAT 578
 QY 181 GluGluTyrValValLeuLysAsnGluMetAlaArgAlaAsnHisTyrGluAspTyr 199
 Db 579 GAAGACTATGTGCTCTGAAAAATGAGATGGCAAGCAACAAATCATATGAGGACTAT 635

RESULT 8
 BI561359
 LOCUS 60325121F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5297380 5',
 DEFINITION mRNA sequence.
 ACCESSION BI561359
 VERSION BI561359.1 GI:15448673
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 609)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabp@remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov
 Plate: LLAM1752 row: 1 column: 05
 High quality sequence stop: 606.
 Location/Qualifiers
 1. 609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5297380"
 /lab_host="NIH_MGC_97"
 /lab_host="DH10B"

/note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site: BamHI; Site: SalI; Site: XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to 500. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 193 a 128 c 137 g 151 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5,29e-93 Length: 609
 Score: 901.00 Matches: 186
 Percent Similarity: 97.40% Conservativity: 1
 Best Local Similarity: 96.88% Mismatches: 2
 Query Match: 21.00% Indels: 4
 DB: 10 Gaps: 0

US-09-635-501-2 (1-805) x BI561359 (1-609)

QY 1 MetSerSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
 Db 40 ATGTCAAGCTCTCTGCTGCTCTCTCAGCCCTGTGCTGTAACTGCTCAGTCCACC 99
 QY 21 IleGluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPhe 40
 Db 100 ATTGAGAACAGGCCAAGACATTTGTGGACAAAGTTTAAACAGCAAGCGGAGACCTGTC 159
 QY 41 TyrGlnSerSerLeuAlaSerTrpAsnTrpAsnThrAsnIleThrGluGluAsnValGln 60
 Db 160 TATCAAGTTCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 219
 QY 61 AsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAl 80
 Db 220 AACATGAATTAATGCTGGGACAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 279
 QY 80 aGlnMetTrpProLeuGlnGluIleGlnAsnLeuThrValLysLeuGlnAlaLe 100
 Db 280 CCAATGTATCCACTACAGAATTCAGAAATCTCAGATCAAGTTCAGTCCAGGCTCT 339
 QY 100 uGlnGln-AsnGlySerSerValLeuSerGluAspLysSerLysArgLeuAsnThr-Ile 119
 Db 340 TCAGCAACAATGGGTCTCTCAGTCTCTCAGAACAGCAAGCAACGGTTGAACACGAAT 399
 QY 120 LeuAsnThrMetSerThrIleTrpSerThrGlyLysValCysAsnProAspAsnProGln 139
 Db 400 CTAATATCAATGAGCACCACATCTACAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459
 QY 140 GluGluLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyrAsn 159
 Db 460 GAATGCTTATTACTTGAACACAGTTTGAATGAATGAATGAATGAATGAATGAATGAAT 519
 QY 160 GluArgLeuTrpAlaTrpGluSerTrpArgSerGluValGlyLysGlnLeuArgProLeu 179
 Db 520 GAGAGCTCTGGGCTTGGGAAAGCTGGAGATCTGAGTGGGAAAG-CAGCTGAGGCCATTA 578
 QY 180 TyrGluGluTyrValValLeuLysAsnGlu 189
 Db 579 TATCAAGACTATGTGCTCTTGAATAATGAG 608

RESULT 9

```

BF789159
LOCUS BF789159 965 bp mRNA linear EST 12-JAN-2001
DEFINITION 602105037F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223312
5', mRNA sequence.
ACCESSION BF789159
VERSION BF789159
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 965)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9811 row: i column: 09
High quality sequence stop: 676.
FEATURES
Location/Qualifiers
source
1..965
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4223312"
/lab_host="NCI_CGAP_Kid14"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 286 a 213 c 253 g 212 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1-88e-88 Length: 965
Score: 864.00 Matches: 182
Percent Similarity: 80.43% Conservative: 16
Best Local Similarity: 73.96% Mismatches: 45
Query Match: 20.14% Indels: 5
DB: 10 Gaps: 2
US-09-635-501-2 (1-805) x BF789159 (1-965)
QY 1 MetSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
Db 21 ATGCCAGCCCTCCCTGGCTCTCTCCAGCCG-TTGTGTTACTACTGCTCAGTCCCTC 79
QY 21 IleGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPhe 40
Db 80 ACCGAGGAAAATCCCAACATTTTAAACAACTTTAATCAGGAAGCTGAGACCTGCTCT 139
QY 41 TyrGlnSerSerLeuAlaSerTrpAsnThrAsnThrAsnIleThrGluGluAsnValGln 60
Db 140 TATCAAAAGTTTCATCTGCTTCTTGGAAATATAATACTAATCACTACTGAGAAAATGCCAA 199
QY 61 AsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAla 80
Db 200 AAGATGAGTGAGCTGCAGCCAAATGGTCTGCTTTTATGAAGAACAGCTCTAGACTGCC 259
QY 81 GlnMetTrpProLeuGlnGluIleGlnAsnLeuThrValLysLeuGlnAlaLeu 100
Db 260 CAAAGTTTCTCATAAGAAAATCCAGACTCCGATCATCAAGCGCTCAACTACAGGCCCTT 319
QY 101 GlnGlnAsnGlySerSerValLeuSerGluAspLysSerLysArgLeuAsnThrIleLeu 120

```

```

Db 320 CAGCAAACTGGGTCTTCAGCACTCTCAGCAGACAGCAAGAAACAGTGAACACAACTTCG 379
QY 121 AsnThrMetSerThrIleTyrSerThrGlyLysValCysAsnProAspAsnProGlnGlu 140
Db 380 AACACCATGAGCACCATTACAGTACTGGAAGAGTTTGCAACCCCAAGAACCCACAAGAA 439
QY 141 CysLeuLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyrAsnGlu 160
Db 440 TGCATTATTACTTGGACGAGGATGGATGAATAATATGGCAGCAGACAGACTACAACTCT 499
QY 161 ArgLeuTrpAlaTrpGluSerTrpArgSerGluValAlaGlyLysGlnLeuArgProLeuTyr 180
Db 500 AGGCTCTGGGCATGGAGGCTGGAGGCTGAGGTGGCAAGCAGCAGCTGAGGCCGTTGTAT 559
QY 181 GluGluTyrValValLeuLeuLysAsnGluMetAla-ArgAlaAsnHisTyrGluAspTyrGln 200
Db 560 GAAGAGTATGTCGTCCTGAAACAGAGATGGCAAGAGCAACAAATTAACGACTATGG 619
QY 200 YAspTyrTrpArgGlyAspTyrGluValAsnGlyValAspGlyTyrAspTyrSerArgGln 220
Db 620 GGATATTGGAGGAGGACTATGAACAGAGGAGGAGCAGATGG-TACAACTATAACCCGTA 678
QY 220 yGlnLeuIleGluAspValGluHisThrPheGluGluIleLysProLeuTyrGluHisLe 240
Db 679 ACAGTGTATTGAGA---TGTGAAGCTACTTCGCAGAAA---TCAGCATTTGGGTGAGCCTCT 732
QY 240 uHisAlaTyrValArg 245
Db 733 TCATGCCCTATGTGAGG 748
RESULT 10
LOCUS AA162058 555 bp mRNA linear EST 12-FEB-1997
DEFINITION ms31b11.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:608613 5' similar to SW-ACE_MOUSE P09470
ANCIOTERIN-CONVERTING ENZYME PRECURSOR, SOMATIC ;, mRNA sequence.
ACCESSION AA162058
VERSION AA162058.1 GI:1739055
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 555)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennan, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:374045
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 446.
FEATURES
Location/Qualifiers
source
1..555
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:608613"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"

```


/lab_host="SOLR (kanamycin resistant)"
 /note="organ: skin; Vector: plasmid SK-; Site: 1: EORI
 ; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo
 dr. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5',
 adaptor sequence: 5' GAATTCGCGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'

BASE COUNT 168 a 106 c 150 g 131 t

Alignment Scores:
 Pred. No.: 5.4e-88 Length: 555
 Score: 857.00 Matches: 155
 Percent Similarity: 89.73% Conservative: 11
 Best Local Similarity: 83.78% Mismatches: 19
 Query Match: 19.97% Indels: 1
 DB: 9 Gaps: 0

US-09-635-501-2 (1-805) x AA162058 (1-555)

QY 126 IletyrSerThrGlyLysValCysAsnProAspAsnProGlnGluCysLeuLeuGlu 145
 Db 1 ATTTCAGTACTGGAAGTTTGCACCCAAAGAACCCACCAAGATGCTATTACTTGAG 60
 QY 146 ProGlyLeuAsnGluMetAlaAsnSerLeuAspTyrAsnGluAtrGLeuTrpAlaTrp 165
 Db 61 CCAGGATTGATGAATAATATGGGCAACACA-GACTACAACTCTAGGCTCTGGGCATGG 119
 QY 166 GluSerTrpArgSerGluValGlyLysGlnLeuArgProLeuTyrGluGluTyrValVal 185
 Db 120 GAGGCTGGAGGCTGAGGTGGCAAGCAGCTGAGGCGCTGTATGAAGATATGTGGTC 179
 QY 186 LeuLysAsnGluMetAlaArgAlaAsnHisTyrGluAspTyrGlyAspTyrTrpArgGly 205
 Db 180 CTGAAAACAGATGCAAGAGCAAAATATATACGACTATGGGATATTGGAGAGG 239
 QY 206 AspTyrGluValAsnGlyValAspGlyTyrAspTyrSerArgGlyGlnLeuLleGluAsp 225
 Db 240 GACTATGACACGAGGACGATGCTGCTACAACTATAACCGTAACCGATGTTGATGAAGAT 299
 QY 226 ValGluHisThrPheGluGluLleLysProLeuTyrGluHisLeuHisAlaTyrValArg 245
 Db 300 GTAGACGTACCTTCGACAGATCAAGCCATTGTATGAGCATCTTCATGCGCTATGTGAGG 359
 QY 246 AlaLysLeuMetAsnAlaTyrProSerTyrIleSerProLleGlyCysLeuProAlaHis 265
 Db 360 AGGAAGTTGATGATACCTACCTCCCTCTACATCAGCCCCACTGGATGCTCCCTGCCAT 419
 QY 266 LeuLeuGlyAspMetTrpGlyArgPheTrpThrAsnLeuTyrSerLeuThrValProPhe 285
 Db 420 TTGCTTGATATGTGGGTAGATTTTGGACAAATCTGTACCTTTGACTGTTCCTTT 479
 QY 286 GlyGlnLysProAsnIleAspValThrAspAlaMetValAspGlnAlaTrpAspAlaGln 305
 Db 480 GCACAAACCAACATAGATGTTACTGATCAATGATCAATGATCAATGATGGGTGGATGCAGAA 539
 QY 306 ArgIlePheLysGlu 310
 Db 540 AGGATATGTCAGAG 554

RESULT 11
 BB652968
 LOCUS BB652968 RIKEN full-length enriched, adult male hippocampus Mus
 DEFINITION musculus cDNA clone C63004JD11 5', mRNA sequence.
 ACCESSION BB652968
 VERSION BB652968.1 GI:15402926
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 686)

AUTHORS

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,H.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)

TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10
 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Yamana,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
 Ishii,Y. and Hayashizaki,Y.
 Mapping of 1932 mouse cDNAs on mouse chromosomes. J. Struct.
 Func. Genomics 2 pre, L72-L86 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

FEATURES source

Location/Qualifiers
 1. .686
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="C63004JD11"
 /clone_lib="RIKEN full-length enriched, adult male
 hippocampus"
 /sex="male"
 /tissue_type="hippocampus"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site: 1: XhoI; Site: 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAAGGATCCAGAGCTCTTTTCTTTTCTTTTCTTTN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adaptor of sequence [5'
 GAGAGAGATTCTCAGTATTAATTAATATCCCTCCCCCCCC 3']"
 211 a 151 c 169 g 151 t 4 others

BASE COUNT ORIGIN

Alignment Scores:
 Pred. No.: 1.74e-83 Length: 586

Score: 819.00 Matches: 162
 Percent Similarity: 81.48% Conservative: 14
 Best Local Similarity: 75.00% Mismatches: 38
 Query Match: 9 Indels: 2
 DB: 9 Gaps: 0

US-09-635-501-2 (1-805) x BB652968 (1-686)

Qy 1 MetSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
 Db 32 ATGTCCAGCTCCCTGCTCCTTCTCAGCCTTGTGTGTTACTACTGCTCAGTCCCTC 91
 Qy 21 lIeGluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspPhe 40
 Db 92 ACCGAGAAATGCGAAGACATTTTAAACAACCTTTAATCAGGAAGCTGGAAGACTGTCT 151
 Qy 41 TyrGlnSerSerLeuAlaSerTrpAsnTyrAsnThrAsnIleThrGluGluAsnValGln 60
 Db 152 TATCAAAAGTTCACCTGCTTCTTGAATTAATACTAATCACTTACGAAGAAATGCCAA 211
 Qy 61 AsnMetAsnAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAla 80
 Db 212 AAGATGAGTGAGGCTGCAGCCAACTGCTGCTTTTATGAAGAACACTCTAAGACTGCC 271
 Qy 81 GlnMetTyrPr-oLeuGlnGluIleGlnAsnLeuThrValLysLeuGlnAlaLeu 100
 Db 272 CAAGTTTCTCAGCTACAAAGAAATCCAGACTCCGATCATCAAGCGTCAACTACAGGCCCTT 331
 Qy 101 GlnGlnAsnGlySerSerValLeuSerGluAspLysSerLysArgLeuAsnThrIleLeu 120
 Db 332 CAGCAAAAGTGGGTCTTCAGCACTCTCAGCAGACAAAGAACACTTGAACACAAATCTG 391
 Qy 121 AsnThrMetSerThrIleTyrSerThrGlyLysValCysAsnProAspAsnProGlnGlu 140
 Db 392 AACACCATGAGCACCACTTACAGTACTGGAAAGTTTGCACCCAAAGACCCACAGAA 451
 Qy 141 CysLeuLeuLeuGluProGlyLeuAsnGluIleMetAlaAsnSerLeuAspTyrAsnGlu 160
 Db 452 TGCCTATTACTTGGACCAAGATTGGATGAATAATATGGCCAGACAGACAGACTACACTCT 511
 Qy 161 ArgLeuTyrAlaTyrGluSerTrpArgSerGluValGlyLysGlnLeuArgProLeuTyr 180
 Db 512 AGGCTCTGGGAGGCTGGAGGCTGAGGTGGTGGCAAGCAGCTGAGGCGCTGTAT 571
 Qy 181 GluGluTyr-ValValLeuLysAsnGluMetAlaArgAlaAsnHisTyrGluAspTyrGl 200
 Db 572 GAAGAGTATGGGNGCTGAAACACGAGAGGCGCAAGAACAAATATAACGACTATGG 631
 Qy 200 YAspTyrTrp-ArgGlyAspTyrGluValAsnGlyValAspGly 214
 Db 632 CGATTATTGGANAGGGGACTTTGAAGCCNAGGGGAGCCAAAGG 675

RESULT 12
 BG428060
 LOCUS 602501471F1 NIH_MGC_75 452 bp mRNA linear EST 14-MAR-2001
 DEFINITION 602501471F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4615121 5',
 mRNA sequence.

ACCESSION BG428060
 VERSION BG428060.1 GI:13334566

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 452)

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LHC1367 row: n column: 18

High quality sequence stop: 450.

FEATURES
 source

Location/Qualifiers

1. 452

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4615121"

/clone_lib="NIH_MGC_75"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggcgctcgcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGCGCCATTATGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.65
 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 144 a 81 c 103 g 124 t

ORIGIN

Alignment Scores:
 Pred. No.: 2,27e-79 Length: 452
 Score: 781.00 Matches: 150
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 18.20% Indels: 0
 DB: 10 Gaps: 0

US-09-635-501-2 (1-805) x BG428060 (1-452)

Qy 570 LeuGluAsnValValGlyAlaLysAsnMetAsnValArgProLeuLeuAsnTyrPheGlu 589
 Db 2 TTGGAATAATGTTGAGGAGCAAGAACATCAATGTAAAGCCACTGCTCAACTACTTTGAG 61
 Qy 590 ProLeuPheThrTrpLeuLysAspGlnAsnLysAsnSerPheValGlyTrpSerThrAsp 609
 Db 62 CCTTATTACCTGGCTGAAAGACCAAGCAAGAAATCTTTTGGGATGGAGTACCGAC 121
 Qy 610 TrpSerProTyrAlaAspGlnSerIleLysValArgIleSerLeuLysSerAlaLeuGly 629
 Db 122 TGGAGTCCATATGCAGACCAAGCATCAAGTGAAGTAAAGCTAAATTCAGCTCTTGA 181
 Qy 630 AspLysAlaTyrGluTyrAsnAspAsnGluMetTyrLeuPheArgSerSerValAlaTyr 649
 Db 182 GATAAGCATATGATGAAGCAGCAATGAATCTACCTGTCGATCATCTGTTGCATAT 241
 Qy 650 AlaMetArgGlnTyrPheLeuLysValLysAsnGlnMetIleLeuPheGlyGluAsp 669
 Db 242 GGTATGAGGAGTACTTTTAAAGTAAATAATCAGATGATTTTGGGGAGGAGGAT 301
 Qy 670 ValArgValAlaAsnLeuLysProArgIleSerPheAsnProPheValThrAlaProLys 689
 Db 302 GTGCGAGTGGCTANTTTGAACCAAGAAATCTCCTTTAAATTTCTTTGTCACCTCACTAAA 361
 Qy 690 AsnValSerAspIleIleProArgThrGluValGluLysAlaIleArgMetSerArgSer 709
 Db 362 AATGTGTCTGATATCATCTCTAGAACTGAAGTTGAAGAGCCATCAGGATGTCCTCGGAGC 421
 Qy 710 ArgIleAsnAspAlaPheArgLeuAsnAsp 719
 Db 422 CGTATCATGATGCTTTCCTCGCTGATGAC 451

RESULT 13

BM030353

LOCUS

DEFINITION 488973 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BM030353

EST 05-NOV-2001

BM030353.1 GI:16743923					
EST.					
Bos taurus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
Bovidae; Bovinae; Bos.					
1 (bases 1 to 556)					
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,					
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett					
,G.L., Heaton,M.P., Laegreide,W.S., Rohrer,G.A., Chitko-McKown,C.G.,					
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and					
Keefe,J.W.					
Sequence evaluation of four pooled-tissue normalized bovine cdna					
libraries and construction of a gene index for cattle					
Genome Res. 11 (4), 626-630 (2001)					
21180013					
Contact: Smith TPL					
USDA, ARS, US Meat Animal Research Center					
PO Box 166, Clay Center, NE 68933-0166, USA					
Tel: 402 762 4366					
Fax: 402 762 4390					
Email: smith@email.marc.usda.gov					
Single pass sequencing. Bases called and alt trimmed with phred					
v0.980904.e. Vector identified by cross_match with the -minscore 18					
and -mismatch 12 options.					
PCR primers					
FORWARD: AGGAACAGCATATGACCAT					
BACKWARD: GTTTCCTCCAGTCAGCAGG					
Plate: 120 row: B column: 16					
Seq primer: ATTAGGTGCACACTATAG.					
FEATURES					
Source					
1..556					
/organism="Bos taurus"					
/db_xref="taxon:9913"					
/clone_lib="MARCB2BOV"					
/tissue_type="pooled"					
/lab_host="DH10B"					
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;					
Library made from pooled tissue from testis, thymus,					
semitendinosus muscle, longissimus muscle, pancreas,					
adrenal, and endometrium."					
BASE COUNT 178 a 111 c 145 g 122 t					
ORIGIN					
Alignment Scores:					
Pred. No.: 1.3e-75 Length: 556					
Score: 749.50 Matches: 142					
Percent Similarity: 86.02% Conservative: 18					
Best Local Similarity: 76.34% Mismatches: 25					
Query Match: 17.47% Indels: 1					
DB: 10 Gaps: 1					
US-09-635-501-2 (1-805) x BM030353 (1-556)					
QY 71 AlaphLeuLysGluInSerThrLeuAlaGlnMetTyProLeuGlnGluLeuAsn 90					
Db 2 GCCTTTTATGAAGAAGCAGTCCCGATGGCCAAACTTACTCACTCGAAGATTGCAAT 61					
QY 91 LeuThrValLysLeuGlnLeuGlnAlaLeuGlnGlnAsnGlySerSerValLeuSerGlu 110					
Db 62 CTCACACTCAAGCGTCAATTGAAGGCCCTTCACATATGAGGGGACCTCACCGCTCGGCA 121					
QY 111 AspLysSerLysArgLeuAsnThrLleLeuAsnThrMetSerThrIleTySerThrGly 130					
Db 122 GAGAAGAGCAACGATTGAACAGGATTCATAAATAAATGAGCACCATCTACAGTACTGG 181					
QY 131 LysValCysAsnProAspAsnProGlnGluCysLeuLeuGluProGlyLeuAsnGlu 150					
Db 182 AAGTTTGAGCCCA--AATACACAGGAGTGTCTAGCACTTGAACACAGGTTTAGATGAC 238					
QY 151 IleMetAlaAsnSerLeuAspTyAsnGluArgLeuTrpAlaTrpGluSerTrpArgSer 170					
Db 151 ILEMETALASNSERLEUASPPTYASNGLUARGLEUTRPALETTPGLUSERTTRPARGSER 170					

```

Best Local Similarity: 85.35%
Query Match: 17.02%
DB: 10
Indels: 13
Gaps: 4

US-09-635-501-2 (1-805) x BG772384 (1-741)

QY 1 MetSerSerSerTrpLeuLeuSerLeuValAlaValThrAlaAlaGlnSerThr 20
DB 58 ATGTCAGCTTCTCCTGGCTCCTCAGCTTGTGCTGAACCTGCTCAGTCCACC 117
QY 21 IleGluGluGlnAlaLysThrPheLeuAspLysPheAsnHisGluAlaGluAspLeuPhe 40
DB 118 ATGAGGACAGCCAGACATTTTGGACAAGTTTAACCAAGACCCGACACCTGTTTC 177
QY 41 TyrGlnSerSerLeuAlaSerTrpAsnTyrAsnThrAsnIleThrGluGluAsnValGln 60
DB 178 TATCAAGTTTCACTTGGCTTCTTGAATATTAACACCAATATTACTGAAGAGAATGCCAA 237
QY 61 AsnMetAsnAlaGlyAspLysTrpSerAlaPheLeuLysGluGlnSerThrLeuAla 80
DB 238 AACATGATATATGCTGGGCAATATGCTGCTTTTATACGGACAGTCCACACTTGC 297
QY 81 Gln-MetTyrProLeuGlnGluLeuGlnAsnLeuThrValLysLeuGlnAlaLe 100
DB 298 CAATATGATCCACTACAAGAAATTCAGAATCTCAGAGTCAAGCTTCAGCTCAGGCTCT 357
QY 100 uGlnGlnAsn--GlySerValLeuSerGluAspLysSerLys-ArgLeuAsnThrI 119
DB 358 TCAGCACAACTGGTCTCTCAGTCTCTCAGAGACAGACACAGGTTGAACACAA 417
QY 119 LeLeu-AsnThr-MetSerThrIleTyrSerThrGly-LysValCysAsnProAsn 138
DB 418 TTCTCAATACATACATGACCACTACAGTACTGGAACCAAGTCTGTAAACCCAGATAATC 477
QY 138 roGlnGluCysLeuLeuGluProGlyLeuAsnGlu-LleMetAlaAsnSerLeuAsp 157
DB 478 CACAGAAATGCTTATTAAGTGTGACCAAGTTTGAATGACACTACTGGCAACAGTTTAC 537
QY 158 TyrAsn---GluArgLeuTrpAlaTrpGluSer-TrpArgSerGluValGlyLysGlnLe 176
DB 538 TACCAACTGACGAGGCTCTGGGCTGGGAAAGCTTGGAGATCT--GCAGGTCGTGCAGC 594
QY 176 uArg-----ProLeuTyrGlu-GluTyrValValLeu 186
DB 595 AGCGTGAGGCCACTTATATGAGGAGTATGCTGGGTCTT 632

RESULT 15
B1826471
LOCUS
DEFINITION
B1826471.1 GI:15938021
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

FEATURES
source
Location/Qualifiers
1..800
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5167987"
/clone_l1b="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 169 a 244 c 230 g 157 t
ORIGIN
Alignment Scores:
Pred. No.: 3.04e-68 Length: 800
Score: 687.50 Matches: 128
Percent Similarity: 68.30% Conservative: 53
Best Local Similarity: 48.30% Mismatches: 81
Query Match: 16.02% Indels: 4
DB: 10 Gaps: 3

US-09-635-501-2 (1-805) x B1826471 (1-800)
QY 230 PheGluGluLysProLeuTyrGluHisLeuHisAlaTyrValArgAlaLysLeuMet 249
DB 9 TTCCAGGAGCTGCAGCCACTCTACTCAACCTGCATGCCTACGTCGCGCGGCGCTGCAC 68
QY 250 AsnAlaTyr---ProSerTyrIleSerProIleGlyCysLeuProAlaHisLeuLeuGly 268
DB 69 CTTCACTAGGGGGCCAGCACACACACCTCAACCTGAGGGGGCCCATTCCTCTCCTCCTG 128
QY 269 AspMetTrpGlyArgPheThrTrpThrAsnLeuTyrSerLeuThrValProPheGlyGlnLys 288
DB 129 AACATGTGGGGCCAGCACCTGCTGCCAACATCTATGACTTGGTGGTCCCTTCCTCCTCAG 188
QY 289 ProAsnIleAspValThrAspAlaMetValAspGlnAlaTrpAspAlaGlnArgIlePhe 308
DB 189 CCTCGATGGACACACACAGGCTATGCTAAACAGGGCTGGAGCCCGCAGGAGGATGTT 248
QY 309 LysGluAlaGluLysPheValSerValGlyLeuProAsnMetThrGlnGlyPheTrp 328
DB 249 AAGGAGGCTGATGATTTCTTCACTCCCTGGGCTGCTGCCCTGCTCCTGAGTTCTGG 308
QY 329 GluAsnSerMetLeuThrAspProGlyAsnValGlnLysAlaValCysHisProThrAla 348
DB 309 AACAGTGCATGCTGGAGAACCCACAGCGGGGGAGGTGCTGCGCGCTGCTCCTGAGTTCTGG 368
QY 349 TrpAspLeuGlyLysGly---AspPheArgIleLeuMetCystThrLysValThrMetAsp 367
DB 369 TGGGACTTCTAACACGGCAGGAGCTTCCGGATCAAGCAGTGCACACCGCTGAACCTGGAG 428
QY 368 AspPheLeuThrAlaHisGluMetGlyHisIleLeuTyrAspMetAlaValAlaAla 387
DB 429 GACCTGGTGGGGCCACACCAAGAAATGGCCACATCCAGTATTTTCATGCAGTACAAAGAC 488
QY 388 GlnProPheLeuLeuArgAsnGlnAlaAsnGluGlyPheHisGluAlaValGlyGluIle 407
DB 489 TTACCTGTGGCTTGGAGGAGGTGTCACCCCGCTTCCATGAGGCCATTTGGGACGTG 548
QY 408 MetSerLeuSerAlaAlaThrProLysHisLeuLysSerIleGlyLeuLeuSerProAsp 427
DB 549 CTAGCCCTCTCAGTGTCTTACCCCAAGCAGCTGCACAGTCTCAACCTGCTGCTGAGCTG 608
QY 428 PheGlnGluAspAsnGluThrGluIleAsnPheLeuLeuLysGlnAlaLeuThrIleVal 447
DB 609 GGTGGCAGCGAC---GAGCATGACATCACTTCTGATGAAGATGCGCCCTTCAACAATC 665

```

```

Qy 448 GlyThrLeuProPheThrTyrMetLeuGluLysTyrArgTyrMetValPheLysGlyGlu 467
Db   ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
666 GCCTTTATCCCTTCAGCTACCTCGTCGATCAGTGGCGCTGGAGGTATTGAT-GGAAGC 724
Qy 468 IleProLysAspGlnTrpMetLysLysTyrTrpGluMetLysArgGluIleValGlyVal 487
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
725 ATCACCAGGAGAACTATAACCCAGGAGTGGTGGAGCCTCAGGCTGAAAAGTACCAGGCCT 784
Qy 488 ValGluProValPro 492
Db   ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
785 CTGCCCCCAGTGCCC 799

```

Search completed: October 9, 2002, 19:18:04
Job time : 1606 secs

